2012

The influence of feedback on self-regulated learning in a Chinese language task

Sarah C. Meacham

THE INFLUENCE OF FEEDBACK ON SELF-REGULATED LEARNING IN A CHINESE LANGUAGE TASK

by

SARAH C. MEACHAM

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of a Masters of Art
in
The Department of Psychology
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

Huntsville, Alabama

2012

In presenting this thesis in partial fulfillment of the requirements for a master's degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department (Director of the Program) or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this thesis.

(student signature) (date) 3/9/12

THESIS APPROVAL FORM

Submitted by Sarah Meacham in partial fulfillment of the requirements for the degree of Masters of Arts in Psychology and accepted on behalf of the Faculty of the School of Graduate Studies by the thesis committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate on the work described in this thesis. We further certify that we have reviewed this thesis manuscript and approve it in partial fulfillment of the requirements for the degree of Masters of Arts in Psychology.

Jodi Price 3/9/12 Committee Chair
(Date)

3/9/12

3/9/12

____

3/9/12 Department Chair

3/12/12 College Dean

Rhonda Kay Gaede 3/15/12 Graduate Dean

ABSTRACT
The School of Graduate Studies
The University of Alabama in Huntsville

Degree Master of Arts College/Dept. Liberal Arts/Psychology

Name of Candidate Sarah C. Meacham

Title The Influence of Feedback on Self-regulated Learning in a Chinese Language Task

Self-regulated learning includes the metacognitive decisions of which items to study, in which order, and how much time to allocate to each item. Performance feedback can be beneficial to the learner and can influence the way individuals allocate their study-time. The present experiment explored how the placement and type of feedback influenced participants' metacognitive awareness, selection of items for restudy, selection order, study-time allocation, and subsequent recall performance. Participants ($N = 183$) received feedback across two trials in a Chinese learning task. The placement of the feedback was manipulated to be item-level, global, or both with half of the participants receiving additional information about the difficulty level of the items and a control group which received no feedback. We expected participants assigned to conditions with more proximal and informative feedback to have improved self-regulated learning. The results provided mixed support for the hypotheses indicating that feedback may have a more intricate and complex relationship with self-regulated learning.

Keywords: Study-time allocation, feedback, item difficulty, metacognition

Abstract Approval: Committee Chair ________________

Department Chair ________________

Graduate Dean ________________ 3/15/12

ACKNOWLEDGMENTS

I would like to thank all of the people who have helped and encouraged me during the work of this thesis. First, I would like to express my gratitude to my advisor, Dr. Jodi Price, whose warmth, guidance, and encouragement throughout all of the stages of the work has not only been invaluable, but also made the process a productive and enjoyable experience. I would like to thank the other members of my committee, Dr. Aurora Torres and Dr. Jeff Neuschatz, who have been very helpful with comments and suggestions as well as supportive in all of my academic endeavors. Also this thesis would not have been possible without the assistance of the lab members who helped collect and record the data. I would like to give a special thanks to Terry Meacham who has supported me not only as my husband, but also as the one who developed the computer program for this project. Lastly, I would like to thank all of my friends and family for always encouraging me and supporting me no matter what challenge I take on.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACROMYMS

ABR: Agenda-based Regulation

ANOVA: Analysis of Variance

DRM: Discrepancy Reduction Model

EOL: Ease of Learning

FOK: Feeling of Knowing

JOL: Judgment of Learning

MCI: Memory Controllability Inventory

RPL: Region of Proximal Learning

LIST OF SYMBOLS

d : Cohen's *d*

F: ANOVA

M: Mean

N: Total sample size

η_p^2: Partial eta squared

p: Significance

SD: Standard deviation

SE: Standard error

t: *t*-test

CHAPTER 1

INTRODUCTION

Metacognition refers to an awareness of one's own cognitive abilities; in other words, it is the knowledge individuals have about their own knowledge (Metcalfe, 2009). This metacognitive awareness is used by the individual to decide how to approach the study of new material in self-regulated learning. When choosing to allocate their study time, people must decide which items to study, the order in which to study them, and how long to spend on each item (Metcalfe & Kornell, 2005). These core decisions are the basis of self-regulated learning behavior. The factors that influence these decisions are of particular interest as they may explain how people self-regulate and why people's study patterns are not always ideal, as well as what can be done to make people more optimal in their self-regulated behavior.

Study-time allocation refers to the specific time one devotes to learning new material and has typically been studied through some variation of a word pair learning task. The experimental stimuli have been mostly English-Spanish vocabulary pairs (Metcalfe, 2002; Metcalfe & Kornell, 2003), Swahili-English vocabulary pairs (Dunlosky & Thiede 1998; Nelson, Dunlosky, Graf, & Narens, 1994), noun-noun word pairs (Ariel, Dunlosky, & Bailey, 2009; Dunlosky & Thiede, 2004; Rhodes & Jacoby, 2007), as well as general information questions (Butler, Karpicke, & Roediger, 2008;

Kornell & Metcalfe, 2006; Nelson & Leonesio, 1988). These experiments have been conducted across all age groups and have encompassed a large range of manipulations. In a typical scenario, participants are presented with to-be-learned material in either a simultaneous (all at once) or sequential (one at a time) presentation format and are given either an unlimited or a constrained amount of time to study these items. Experimenters often give participants different objectives, such as speed, accuracy, or particular learning goals (e.g., learn 6 of the 30 items). The encoding phase is followed by a cued recall test on the material. Some experiments also include multiple study-test trials over the same material to observe reselection behaviors (Son & Metcalfe, 2000).

Metacognitive Judgments

To observe the metacognitive processes that occur when participants allocate their study time in a self-regulated learning task, various metamemory judgments are often collected throughout the experiment from the participants. One judgment, called an ease of learning (EOL) judgment, occurs prior to the learning phase. Typically, participants will view half of a to-be-learned word pair and make a judgment as to how easy it will be to learn when presented as a pair. For example, if using Spanish-English word pairs, the participant may see a Spanish word and rate how easy it would be for them to learn the English equivalent without actually knowing what the English equivalent is. Another judgment is called a feeling of knowing (FOK) judgment. This judgment is made during or after the learning has occurred to see whether the item is known or will be remembered on a later test when it is currently unable to be recalled. A judgment of learning (JOL) is also collected during or after the learning phase, but it is a prediction about future test performance on currently recallable items (Nelson & Narens, 1990). For example, after

studying a complete word pair, participants might be asked to rate how likely they would be to remember it on a subsequent recall test and this would be their JOL.

These metacognitive, self-judgments are usually compared to the actual recall accuracy in testing to see how aware people are of their learning progress. There are generally two ways these comparisons can be made. There are absolute comparisons which are based on the mean difference between the predicted and actual performance. There are also relative comparisons which are calculated with one item relative to another with gamma correlations. These are based on the item level differences between predicted and actual performance which take into account whether items given higher ratings are in fact more likely to be recalled than items given lower ratings (Nelson, 1984; Nelson & Dunlosky, 1991). Relative accuracy, calculated through a Goodman-Kruskal gamma correlation, has the most advantages for comparing metacognitive judgments and recall performance in a self-regulated learning task because gamma coefficients are not affected by mean level differences in ratings the way absolute accuracy is (Nelson, 1984).

Research has shown a direct relationship between these judgments and the allocation of study time (Nelson & Leonesio, 1988). This relationship is not always optimal. Poorer accuracy-judgment correlations are usually associated with less optimal study time allocation (Metcalfe & Kornell, 2003). JOLs are particularly useful in understanding the metacognitive processes involved in self-regulated learning. In several experiments where the restudy of items was manipulated across multiple trials, items with lower JOLs were more often selected for restudy. This is optimal for the learning process because people are selecting items that they do not already know. This indicates that people are generally metacognitively aware of how well they have learned an item,

which offers important insight into the learning process (Nelson et al., 1994). Unfortunately, when JOLs do not match true learning, people still rely on them when making study choices. For example, Soderstrom and McCabe (2011) found that people give higher JOLs for information they deem as more important, such as information worth more points, but that these JOLs can be erroneous. Incorrect monitoring of JOLs can lead to poor self-regulated learning and task performance (Metcalfe & Finn, 2008).

The timing and presentation of JOLs affect their accuracy. JOLs can either be made immediately after studying an item or after a delay. Delayed JOLs have been shown to be more accurate predictors of recall performance than immediate JOLs. When participants provided JOLs shortly after study instead of immediately after each word the delayed judgments more precisely reflected participants' recall performance (Nelson & Dunlosky, 1991). Another factor that influences the accuracy of JOLs is how much information participants are given when providing their JOLs (i.e., whether JOLs are based on cue-target pairs or cues only). JOLs have been found to be more accurate when the participant is presented with only the cue (e.g., only the Spanish word) instead of the intact cue-target (e.g., Spanish – English) stimulus pair (Dunlosky & Nelson, 1997). Another influence over the accuracy of JOLs is the amount of trials to which participants have been exposed. As the number of trials increases, the accuracy of the JOLs increases (Koriat, 1997). When these factors are taken into consideration, JOLs can be very useful in predicting recall performance and understanding a person's metacognitive control over self-regulated learning.

A JOL is based on individuals' perceptions of how likely they will be able to recall an item on a future recall test (Nelson & Narens, 1990). One factor that influences

participants' JOLs is item difficulty (Dunlosky & Matvey, 2001; Koriat, 1997). Item difficulty may be manipulated in a variety of ways, including whether word pairs are related or unrelated (Dunlosky & Matvey, 2001), concrete or abstract (Pelegrina, Bajo, & Justicia, 2000), or whether foreign language vocabulary terms resemble their English equivalents (i.e., are cognates) or not (Metcalfe, 2002; Price, Hertzog, & Dunlosky, 2010). Thus participants tend to give lower JOLs to unrelated, abstract, and non-cognate items than to related, concrete, cognates, which reflects less confidence in their ability to recall these items on an upcoming test. The fact that participants are sensitive to these manipulations when providing JOLs is important because there is a strong correlation between JOLs and the allocation of study time (Metcalfe & Kornell, 2003).

There has been much debate on the role that an item's perceived level of difficulty has on a person's self-regulated study behavior. This debate has yielded three prominent models with different predictions as to what people focus on when making self-regulated learning decisions and what influences these decisions.

Discrepancy Reduction Model

The discrepancy reduction model (DRM), was first introduced by Dunlosky and Hertzog in 1998. They suggested that people allocate more study time to items that are perceived to be more difficult to learn because it takes more time to reduce the discrepancy between their current state of learning and their overall goal of learning for these items. This theory is based on the norm-affects-allocation hypothesis, which states that individuals will change their desired degree of learning, referred to as the norm of study, in order to achieve a task goal (Nelson & Leonesio, 1988). When the norm of study is increased more time will be allotted to the item's study (Dunlosky & Hertzog,

1998). Support for the DRM has been most often seen when there is an unlimited amount of time for learning the items, the goal has been complete mastery, and when the person studying the materials is an expert in the area (Metcalfe, 2002).

The DRM has much empirical support (Son & Metcalfe, 2000), but the conclusions drawn from this research may be limited in their scope and may lack the ecological validity needed to understand learning processes. Metcalfe (2002) suggests that many of the experiments that support the DRM are limited because they had no time constraints and mastery was emphasized, but that DRM-consistent study behaviors only make sense when people have unlimited time to study. Moreover, the items were presented one at a time and reselections were not allowed (Metcalfe, 2002). These experiments have also been criticized for the materials used for study. It has been hypothesized that the level of item difficulty was not salient enough to alter a person's selection choice (Metcalfe, 2002). For these reasons the DRM has recently fallen out of favor and research regarding more natural learning situations has been conducted.

Region of Proximal Learning Model

The region of proximal learning model, or RPL, was first introduced by Metcalfe in 2002. She argues that people allocate more study time to items that are perceived as easier to learn because the easier items are closest to participants' current state of knowledge. This is largely based on the transitional state in Vygotsky's zone of proximal development model (Metcalfe, 2002). RPL-consistent study behaviors have been observed most often when there is a limited amount of time, the goal of studying is less than mastery, and when the person studying the materials is a novice in the area. There has been much empirical support for RPL with many experiments showing that people

tend to focus on easier unknown items when they are novices and have a limited amount of study time (Metcalfe & Kornell, 2003, 2005; Price, Hertzog, & Dunlosky, 2010).

Agenda-based Regulation Model

Both the DRM and RPL model concentrate on the role of item difficulty in item selection and study time allocation. DRM predicts that people will select items that are more difficult for study while RPL predicts people will select easier items for study. An explanation for why both selection patterns have been observed is that the goals motivating the person's decisions in self-regulated study are actually influencing the selection strategy, not just the item's level of difficulty (Dunlosky, Ariel, & Thiede, 2011). Agenda-based regulation of learning, or ABR, was first introduced by Ariel et al. (2009) to address the conflicting research regarding RPL and DRM. They argue that people allocate their study time based on a combination of goal based and stimulus driven responses. When there are goals present, such as point values or the likelihood of the material being on a test, top-down processing can influence study-time allocation choices because people are able to create an agenda based on these goals to direct their self-regulated learning. With these goals, the impact of item difficulty on item selection can be overridden to make individuals spend more time on easier or more difficult items, depending on which one benefits them the most for the test. When there are no goals to affect the decision, bottom-up, stimulus driven factors, such as item difficulty, will influence how individuals choose to allocate their study time (Dunlosky et al., 2011).

The ABR model also provides an account for habitual selection and study time allocation behaviors. ABR suggests that individuals self-regulated learning behaviors may reflect habitual processes, such as selecting items in a left-to-right reading order, that

occur without regard to item difficulty. Dunlosky and Ariel (2011) found a left-to-right selection pattern for native English speakers. The opposite pattern has been observed in readers whose native language is written in a right-to-left order (Ariel, Al-Harthy, Was, & Dunlosky, 2011). These experiments showed how RPL can be overridden by habitual responses (i.e., reading order) and provided support for the ABR model by demonstrating that selection decisions can be influenced by something other than item difficulty.

Individual Differences

The learning agenda set by an individual during a learning task and thus the elements of the task focused on to reach this agenda (e.g. item-difficulty, point values, reading order, etc.) may depend on the memory self-efficacy and the working memory span of the individual. Memory self-efficacy refers to the beliefs held by an individual regarding their own memory abilities. Memory self-efficacy can influence the individual's persistence during the task, the amount of effort put forth, the goals set, the strategies used, and the choice of what to study. This influence over self-regulated learning behaviors can affect how well the individual ultimately performs on a task (Berry & West, 2002). This relationship may also be affected by the complexity of the task. An individual's memory self-efficacy tends to have greater influence over self-regulated behaviors in tasks which are more cognitively demanding (Beaudoin & Desrichard, 2011).

Previous research has also indicated an association between people's working memory span and how they choose to allocate their study time in self-regulated tasks (Dunlosky & Thiede, 2004; Price et al., 2010). Those with lower working memory spans may have more difficulty with their attentional control during complex tasks. This may

influence the goals individuals set for themselves as well as their ability to successfully implement strategies to achieve these goals (Dunlosky & Thiede, 2004). Thus individual differences among participants in memory self-efficacy and working memory span can influence self-regulated behavior.

Feedback

A major influence over the way people approach self-regulated tasks can be the feedback they receive on their performance. Feedback is deeply intertwined with self-regulated learning. It gives external information about individuals' cognitive engagement which has a direct effect on their level of achievement (Butler & Winne, 1995). Feedback allows individuals to compare their metacognitive understanding of how well they perform to their actual performance, thus informing them of any discrepancies. By addressing these discrepancies, there can be an increased amount of metacognitive control during any subsequent, related tasks. It is not surprising that feedback has been shown to influence the allocation of study time (Rhodes & Jacoby, 2007) and to be very beneficial in learning (Metcalfe & Kornell, 2007).

From an educational standpoint, feedback has many useful applications. Nicol and Macfarlane-Dick (2006) outline seven principles for good feedback practices in self-regulated learning. The principles reflect the valuable impact that feedback may have on learning when feedback is used optimally. They offer that feedback can clarify what good performance is to the learner by giving information on the goals, criteria and standards expected and in turn enhance performance. Feedback can guide behavior by offering effective information to help close the gap between the current and desired state of learning. Feedback guides self-assessment, which can aid in the development of better

monitoring and more effective study strategies. Good external feedback can help students identify errors in their own performance so they can successfully correct themselves. In a classroom setting, feedback can encourage communication between student and teacher to facilitate learning. Feedback is also useful for the teacher because it helps them understand when the students are in error so they can mold their teaching accordingly. Internally, feedback can influence motivation and increase self-esteem when administered properly. Feedback can particularly be beneficial in low-stake situations, such as situations where feedback focuses on general progress instead of success or failure. These principles reflect the usefulness of feedback in an educational setting and underscore the importance of understanding the optimal way to give feedback.

Feedback has been investigated in motor skill learning and performance. Chiviacowsky and Wulf (2007) found feedback to be more beneficial when a person initially did well on a task than when they did poorly. Performance was also enhanced when the person was able to control when they would receive the feedback than when it was experimentally controlled. Wulf, Chiviacowsky, and Lewthwaite (2010) found motivational influences of normative feedback. In their experiment, they gave two groups differing, arbitrary feedback regarding their group's performance compared to the average performance in a motor task. One group received feedback that they did worse than average and the other group that they did better than average. The group with the positive feedback performed better the next day on a similar task. The influence of positive feedback has also been observed in metacognitive studies. West, Welch, and Thorn (2001) found that when participants were given feedback that was positive, their

performance, motivation, and memory self-efficacy increased, but when they were given negative feedback the reverse occurred. Since the feedback that was given in this experiment was based on actual performance, this was particularly detrimental to older adults and led to very poor recall performance.

Although negative feedback can be detrimental to motivation and performance, feedback which simply corrects incorrect responses can help subsequent recall performance. A hypercorrection effect has been observed with high-confidence incorrect responses when feedback is given to the correct response. This means that individuals are actually more likely to get an incorrect item correct on a subsequent test when they have higher confidence in their incorrect answer when given corrective feedback. Retention in subsequent tasks also is greater for these high-confidence errors than other errors (Butterfield & Metcalfe, 2001; Metcalfe & Kornell, 2007). Low-confidence correct responses have been shown to benefit from feedback as well, with retention nearly doubling with the addition of feedback (Butler et al., 2008). Generally, corrective feedback has a positive influence over later performance in self-regulated tasks (Metcalfe & Kornell, 2007).

The timing when feedback is given can influence whether the feedback benefits the learner. Delayed feedback, or feedback given after a given amount of time has passed, has been shown to be more beneficial with relatively simple tasks and has been associated with better transfer of learning. Immediate feedback, or feedback given directly after a response or at the end of a task, has been shown to be more beneficial when a task is more difficult and has been associated with more immediate gains in

learning (Shute, 2008). Although it depends on the type of task, generally immediate feedback has been found to be more beneficial than delayed feedback for learning (Kulik & Kulik, 1988). Students also perceive immediate feedback to be more useful than delayed feedback, particularly when the feedback includes additional information about the correct answer (van der Kleij, Eggen, Timmers, & Veldkamp, 2012).

Studies have shown that immediate feedback, when presented cumulatively at the end of a recall test, has a smaller influence than feedback following each response (Kulik & Kulik, 1988). Ultimately feedback which corrects the incorrect responses has been shown to be the most effective and is more powerful than just giving feedback as to whether the response is correct or incorrect. Interestingly, any information provided in addition to the correct answer has largely been found to have no further impact on later recall (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Although additional information generally does not affect recall, elaborated feedback may help reduce the cognitive load of a task, which can help later performance in some tasks (Moreno, 2004). Benefits of immediate corrective feedback have also been demonstrated in a classroom setting. Dihoff, Brosvic, and Epstein (2003) found an increase in confidence and accuracy in the responses on an end of semester test when students were given immediate item-level feedback during their quizzes throughout the semester compared to a group who received immediate overall feedback (i.e., information about how many items total they had answered correctly) after each quiz or after a 24 hour delay.

Meta-analyses have shown the benefit of immediate feedback over delayed feedback across different instructions and circumstances (Bangert-Drowns et al., 1991; Kulik & Kulik, 1988). Yet the mechanisms behind these observations are not fully

known. A central issue is how to construct feedback to be the most useful and beneficial to the learner. This would include where the feedback should be placed and the information the feedback provides. The present experiment investigated whether the placement of feedback in a two trial self-regulated learning task influences reselection strategies and recall performance. Of particular interest was how information about item difficulty, coupled with feedback, affected participants' later selection of items to restudy and recall performance.

Hypotheses

We were interested in how the placement and informativeness of feedback influenced participants' self-regulated learning in a Chinese learning task. In particular, we were interested in feedback's influence over the metacognitive awareness of participants, which items they selected for restudy, the order in which participants selected items for restudy, the amount of time spent on those items, and participants' subsequent recall performance. We manipulated when participants received feedback (global, item-level, or combination) and whether or not they received information about how difficult the items were. We expected there to be differences in self-regulated behaviors in the second trial based on the assigned condition. The control condition, where the participants received no feedback, served as a baseline to which the other conditions were compared. Having this baseline control condition allowed us to not only determine if feedback improved performance, but also whether it hindered performance based on the type.

JOLs. By analyzing the relationship between JOLs and recall test performance we hoped to be able to determine how feedback influences a person's metacognitive

monitoring. The more feedback given to people the more they should be able to fine tune their metacognitive monitoring and in turn make more accurate predictions about their recall performance. As the specificity and the amount of the feedback increases, the concordance between JOLs and recall accuracy should increase as well. Consequently we expected that participants assigned to conditions which had more proximal and informative feedback would have greater JOL accuracy than participants in conditions with more distal and less informative feedback. Specifically, in terms of JOL accuracy, we expected the control condition should be lowest, the global feedback conditions should be low, the item-level feedback conditions should be moderate, and the combination feedback conditions should be highest. Conditions that included difficulty information should have higher correspondence between JOLs and recall than those without this information. The combination condition which utilizes feedback at both the global and item-level and includes difficulty information was predicted to have the highest relative accuracy of JOLs. There should also be a general increase in the relative accuracy of JOLs from trial 1 to trial 2 as the accuracy tends to increase with practice (Koriat, 1997). Similar expectations existed for the absolute accuracy of participants JOLs across trials.

Selection. During the reselection phase, we expected participants to select easier items which were not recalled on the trial 1 recall test. In conditions with item-level feedback participants should have been more aware of which items they specifically got correct or incorrect and for that reason we expected that they would be more accurate in their reselections. In addition, we expected that conditions with difficulty information would serve to help participants manage any cognitive overload by allowing them to

generalize which type of items they should concentrate on if they could not remember precisely which items they got correct or incorrect. Thus, we expected participants to concentrate on the main type(s) of items they got incorrect during the reselection phase. If the participant got most of the simple items correct we expected them to concentrate on the more complex items, but if they missed a lot of the simple items we predicted that they would concentrate on those.

In sum, participants were expected to try to study within their own region of proximal learning. After participants selected half of the items for restudy, we expected participants to initially select simple items and to gradually move to more complex items during the second trial's study phase, in accordance with the RPL model (Metcalfe, 2002). We also expected participants to spend proportionally more time on difficult items in accordance with the DRM. In addition to the influence of item difficulty, we expected for grid item selection to generally occur in a left-to-right selection order similar to Western reading as predicted by the ABR model (Dunlosky & Ariel, 2011).

Study-time. Participants were expected to initially spend time on easier items, but to proportionally spend more time on the difficult items, in accordance with the DRM (Dunlosky & Thiede, 1998). Additional information about the differences between the characters was expected to help participants optimize study-time allocation by making the participants more aware of what they needed to concentrate on (Pelegrina et al., 2000). In addition, we expected item-level feedback to help participants remember exactly which of the 36 characters were correct or incorrect and thus help them allocate their study-time appropriately.

Recall performance. For trial 1, it was expected that recall performance would be similar across all conditions given that all feedback manipulations initially occurred during the Trial 1 recall test. We expected participants in all conditions to show improved recall performance in trial 2 given the opportunity to restudy many of the items again. However, we expected trial 2 to reveal recall differences based on the assigned condition. Participants in the combination conditions, which contained both item-level and global feedback, were expected to achieve the highest level of recall because they were expected to have the highest metacognitive awareness and to be the most optimal with their reselections since they were given the most information about their performance. Participants in the item-level conditions were expected to have the next highest rates of recall in trial 2, followed by those in the global feedback conditions. In all cases, conditions which included item difficulty information were expected to outperform those without this information based on the assumption that this would inform participants about which items needed to be studied after the first recall test. Finally, we expected participants to recall more of the simple items than the complex items in accordance with the RPL model (Metcalfe, 2002).

CHAPTER II

METHOD

Participants

Undergraduate students from The University of Alabama in Huntsville (*N* = 183) participated in exchange for credit in their introductory psychology classes. The average age was 21.23 (*SD* = 4.94), and 61.5% of the participants were female. Participants were excluded if they were familiar with Korean, Japanese, and Chinese languages to insure that all participants were novices. Participants under the age of 19 were required to have parental consent and all APA ethical guidelines were followed (see Appendix A for UHSC form).

Random assignment of participants to conditions yielded 26 in the control condition, 25 in the global with difficulty information condition, 26 in the global without difficulty information condition, 28 in the item-level with difficulty information condition, 26 in the item-level without difficulty information condition, 25 in the combination with difficulty information condition, and 27 in the combination without difficulty information condition. Thus all conditions had similar numbers of participants.

Design

The experiment used a 2 (Trial: 1 or 2) x 3 (Item type: simple, moderate, or complex) x 3 (Feedback type: item-level, global, or combination) x 2 (Difficulty

information: present or absent) mixed factorial design. Trial and item type were manipulated within subjects and feedback and difficulty information were manipulated between subjects. There was also a control group where there was no feedback or difficulty information. A trial consisted of one study session and one recall test. Item type referred to the normative difficulty of the Chinese characters, which was determined based on prior research using the same stimuli (Price & Murray, under review). Feedback type referred to the type of feedback the participants received during their recall tests, whether it was item-level (after each recall response), global (at the end of the recall test), or a combination of the two. Item difficulty referred to whether the participants received difficulty information along with their feedback (i.e., information about whether items were simple, moderate, or complex).

Materials

Demographics. Basic demographic information such as age, sex, and race was collected prior to the experiment through a mostly multiple choice, paper-based questionnaire, which is presented in Appendix B.

Memory Controllability Inventory. Participants' beliefs about their own memory were assessed with the Memory Controllability Inventory (MCI; Lachman, Bandura, Weaver, & Elliott, 1995). The MCI was used to observe whether people's beliefs about their memory had any connection to how they performed in the Chinese learning task (e.g., their willingness to select more complex items). This paper-based questionnaire consisted of 23 statements where participants were asked to make a response using a 1 to 7 Likert scale (1 = strongly disagree; 7 = strongly agree) with some

of the statements reversed scored. These statements addressed memory beliefs about the self, age-related memory beliefs, and beliefs about Alzheimer's disease.

Listening Span Task. To assess participants' working memory capacity the Listening Span Task (Salthouse & Babcock, 1991) was used and presented with an audio recording. This task required participants to listen to a simple sentence and answer a multiple choice question about that sentence. The participants then had to write what the last word of the sentence was at the end of each trial. There were 3 trials in each section and there were 7 sections total. Each section added another sentence to the trials until there were 7 sentences for which the last words had to be recalled at the end of each trial. We included the Listening Span Task to examine whether a person's working memory span influenced our results.

Chinese Learning Task. The stimuli consisted of 36 Chinese characters varying in difficulty level. Price and Murray (under review) assessed difficulty level and found that as the number of brush strokes necessary to draw each character increased so did the perceived level of difficulty. The 36 characters consisted of 12 Simple characters (2 - 4 brush strokes), 12 Moderately complex characters (5 - 9 strokes), and 12 Complex characters (10 - 14 strokes). All of the Chinese characters used in the program and their normative difficulty levels are presented in Appendix C. The Chinese task was a computer-based program which was created within the laboratory.

Post-task Questionnaire. After the experiment was completed, a post-task questionnaire was administered. This questionnaire was created in house and was tailored to address the manipulations of each condition. This allowed us to collect information regarding the strategies participants reported using in selecting items for

study and whether these reports differed as a function of the various feedback conditions. An example of this questionnaire, which includes all of the possible questions which were used, is presented in Appendix D.

Procedure

After proper consent was obtained, participants were randomly assigned to one of the feedback conditions and seated in front of a computer with up to 8 participants per session. Participants in each session received the same condition as one another to prevent diffusion of treatment.

Participants were first asked to provide basic demographic information followed by the MCI (Lachman et al., 1995). These tasks were self-paced and took approximately 10 min. After these were completed, the Listening Span Task (Salthouse & Babcock, 1991) was administered via audio recording and generally took about 30 min to complete. Participants were then instructed to begin the self-paced Chinese learning task computer program which took about 30 min to complete, but this varied from participant to participant.

In the first phase of the computerized Chinese learning task, participants were asked to give EOL judgments about how easy or difficult each of the 36 Chinese characters would be to learn. These ratings were made on a 1 to 9 Likert scale, with 1 being easiest to learn and 9 being the hardest. In this phase they were not able to see the English equivalent, only the Chinese character. After EOLs were obtained, participants were asked to make a prestudy prediction of how many of the 36 characters they expected to be able to recall after being given 3 s to study each word pair.

Participants then began the first study trial which showed each of the Chinese characters with their English equivalent for 3 s. After all of the 36 characters were studied, delayed JOLs were collected for each of the items in a newly randomized order. Participants only saw the Chinese character on the screen and were asked to provide a JOL on a scale of 0 to 100, with 0 being a judgment that they would be 0% likely to recall the item on a recall test and 100 being they were 100% certain they would recall the English equivalent for the presented Chinese character. After the 36 characters were studied and JOLs were made, participants were asked to provide a post-study prediction of how many of the 36 characters they expected to be able to recall on the recall test.

This was followed by the actual cued-recall test. The participant was presented with a Chinese character and asked to type the English equivalent. If a participant did not know the English equivalent, they were allowed to leave the answer blank and were instructed to hit the "Continue" button to move on to the next character. In keeping with prior research, the items were considered correct based on the first three letters of the word to avoid spelling or typing errors and to insure more accurate feedback (Nelson et al., 1994). None of the 36 Chinese characters had the same first three letters in order to avoid any false feedback during the recall test as to whether an item was correct or incorrect.

For the item-level feedback conditions, the recall test consisted of participants typing their response and then hitting the continue button. This was followed by a feedback screen simply indicating whether the answer was correct or not, but not revealing the correct answer, such as, "You got the previous answer correct." During this screen the participant was not able to see the word pair. The feedback message was

displayed until participants hit continue to view the next character on the recall test. This process continued until all 36 characters had been tested. In the item-level condition which included normative difficulty information, participants were also given information about whether the item was simple, moderate, or complex such as, "You got the previous answer incorrect, which was a Moderate character."

For the condition with global feedback, the cumulative feedback was displayed at the end of the recall test, such as, "You got 10 items correct this trial." In the global feedback condition which included normative difficulty information participants additionally received cumulative feedback about the difficulty of the characters, such as, "You got 10 items correct this trial (Simple: 6; Moderate: 4; Complex: 0)."

In the condition with the combination feedback, the participants received feedback information both at the item and global level. In the combination condition with difficulty information, the normative item difficulty was also given at both the item and global level. Therefore, participants in this condition were exposed to feedback multiple times with item level feedback, "You got the previous answer incorrect, which was a Simple character," as well as global feedback at the end of the recall test, "You got 15 items correct this trial (Simple: 8; Moderate: 4; Complex: 3)."

In the control condition, participants received no feedback of any kind. After the recall test was completed, participants were asked to postdict how many of the 36 characters they think they recalled correctly. Participants in conditions with global feedback postdicted their recall test performance after the recall test but before receiving any feedback. Participants in conditions with item-level feedback postdicted their performance at the end of the recall test.

After this first recall test was completed, participants then studied for a second recall test. First they were asked to make a prestudy prediction as to how many of the 36 characters they expected to recall correctly on a second recall test when given an opportunity to restudy half of the items. Participants then saw all 36 Chinese characters in a 4 x 9 grid and were instructed to select 18 of the items for restudy by checking a box which was displayed under each of the characters. The 18 items participants selected were randomly placed into a 3 x 6 grid and participants were given 45 s of study time to allocate however they chose among the 18 items. The participants were given 45 s to allocate based on a previous experiment in the laboratory which found 30 s to study 9 Chinese characters generally allowed participants to study every character with extra time. Therefore the time was cut slightly to 45 s for 18 items to increase the likelihood that participants would experience some time pressure.

Once this second study phase was complete, a second set of delayed JOLs was collected over all of the 36 characters in the same manner as the first trial. Participants were then asked to make another poststudy prediction as to how many of the 36 characters they expected to recall on the recall test. This was followed by a second recall test over all of the characters with the same condition specific feedback as they received in the first trial. They were also asked to give another postdiction of their performance just as in the first trial for all 36 characters. In all phases of the program, characters were randomly presented.

Following the computerized Chinese task, participants were asked to complete the post-task questionnaire. They were then debriefed, assigned activity points for their

classes, and dismissed. The entire experiment lasted between 60 and 80 min depending on the pace of the participants in each session.

Statistics

The data were converted from EXCEL to SPSS using SAS, which condensed the lines of data to a usable format for analysis. Goodman-Kruskal gamma correlations were calculated to assess the relative accuracy of the JOLs. As this experiment utilized two trials, repeated measures analyses of variance (ANOVAs) were used to compare the relative and absolute accuracy of the JOLs and recall performance across trials. Conditional probabilities were calculated to see whether a participant's performance in trial 1 influenced their selection pattern in trial 2. Of particular interest was whether the nature and timing of feedback affected how optimal participants' reselections were. For example, it would be optimal for someone who did well in the first trial and recalled most of the simple and moderate items, but missed several complex items, to select more of the complex items for restudy than a participant who missed more of the easier and moderate items for which it would make more sense to reselect these easier items. The analyses focused on whether some types of feedback yielded more optimal reselections than others, which should have positively affected the accuracy of JOLs and participants' recall performance.

CHAPTER III

RESULTS

Metamemory Judgments

EOLs. Given that participants' perceptions of item difficulty could be expected to influence their self-regulated learning decisions, we began by examining participants' EOLs, which were collected for all 36 of the Chinese characters prior to the study phase of the experiment. These judgments were based on a 1 to 9 Likert scale with an increase in number indicating an increase in perceived difficulty. A 3 x 7 repeated measures ANOVA was conducted with normative Item difficulty level (simple, moderate, or complex) as a within subjects factor and Condition as a between subjects factor. Of interest was whether participants' initial perceptions of item difficulty differed as a function of either normative difficulty level or condition, which could affect participants' self-regulated learning decisions. There was a significant main effect for Item difficulty level, $F(2, 175) = 9.15$, $p < .01$, $\eta_p^2 = .91$, because participants' EOL ratings increased as the normative complexity level of items increased (see Figure 3.1). There was no significant interaction between perceived Item difficulty level and Condition, $F(12, 352) = 1.17$, $p = .31$. This was expected because the EOL ratings were made prior to any of the feedback manipulations and this pattern has been seen in previous research (Price et al., 2010).

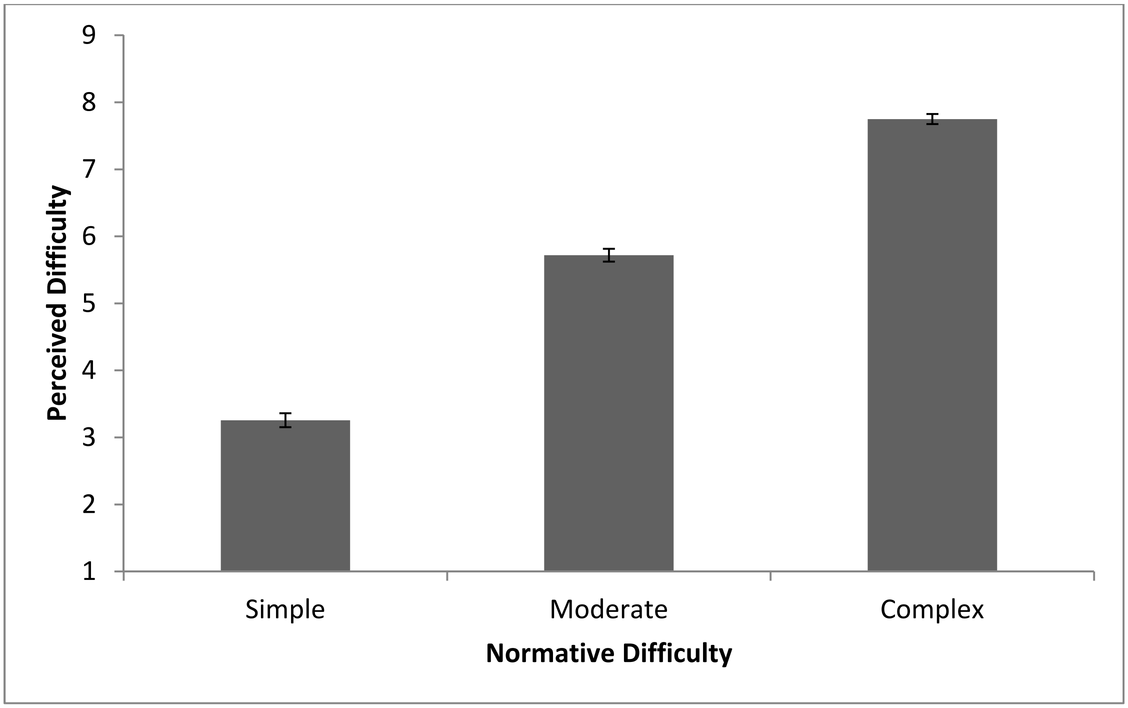

Figure 3.1: EOL Ratings for Normatively Simple, Moderate, and Complex Items.
Note. Ratings of 1 indicate the item is perceived as easy to learn and 9 indicate the items is perceived as difficult to learn.

Mean predictions and postdictions. Global predictions and postdictions were collected throughout each trial. The participant was asked to predict how many of the 36 items they expected to be able to recall correctly on a subsequent recall test before and after each study phase. The participant was also asked to make a postdiction as to how many they believed they had gotten correct after each recall test. A 2 x 3 x 7 ANOVA was conducted with Trial (1 or 2) and Judgment type (prestudy prediction, poststudy prediction, or postdiction) within subjects and Condition between subjects. The means and standard errors for Trial and Judgment type are reported in Figure 3.2. There were significant main effects for Trial, $F(1, 176) = 113.90$, $p < .01$, $\eta_p^2 = .39$, and Judgment type $F(2, 175) = 40.13$, $p < .01$, $\eta_p^2 = .31$. There was also a significant interaction between Trial and Judgment type $F(2, 175) = 71.70$, $p < .01$, $\eta_p^2 = .45$. The interaction

was driven by the fact that in the first trial, the prestudy predictions were higher than in the second trial, but once the participants were given a chance to see the Chinese characters their ratings were much lower for the poststudy predictions and the postdictions. In trial 2 the poststudy prediction and postdictions were significantly higher than those given in trial 1. This was expected since the participants saw the characters for the second time in trial 2 and should have been able to recall more items after a second study session.

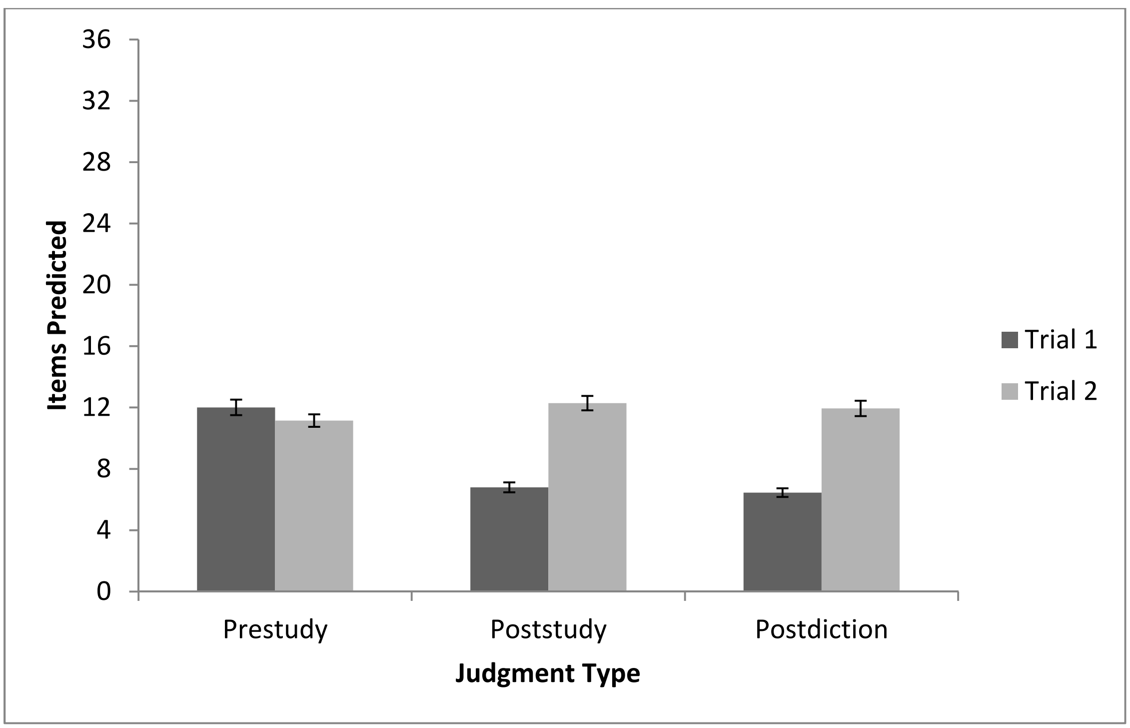

Figure 3.2: Predictions and Postdictions as a Function of Trial.

There was also a significant interaction between Judgment type and Condition, $F(12, 352) = 2.04$, $p = .02$, $\eta_p^2 = .07$, and for Trial and Condition $F(6, 176) = 2.65$, $p = .02$, $\eta_p^2 = .83$. The means and standard errors for Judgment type as a function of Condition and Trial appear in Table 3.1. In trial 1 prestudy predictions were generally the highest, followed by poststudy predictions, and then postdictions were the lowest. In trial 2 the prestudy predictions, poststudy predictions, and postdictions varied as a function of Condition, but generally participants' expectations for their performance were higher in trial 2 than in trial 1. Participants in conditions which included difficulty information generally gave lower predictions and postdictions after receiving the feedback than participants in conditions which did not include difficulty information.

Table 3.1

Predictions and Postdictions as a Function of Condition and Trial

		Prestudy	Poststudy	Postdiction
Condition	Trial	Mean (*SE*)	Mean (*SE*)	Mean (*SE*)
Control	1	10.08 (1.35)	6.54 (.85)	6.69 (.73)
	2	11.58 (1.10)	12.50 (1.25)	12.89 (1.31)
Item-level w/Diff	1	12.96 (1.37)	7.08 (.86)	6.32 (.74)
	2	10.20 (1.12)	11.16 (1.23)	11.12 (1.34)
Item-level	1	13.77 (1.35)	7.73 (.85)	8.12 (.73)
	2	14.00 (1.10)	13.46 (1.25)	14.27 (1.31)
Global w/Diff	1	12.75 (1.30)	6.11 (.81)	4.71 (.70)
	2	8.75 (1.06)	9.79 (1.21)	7.86 (1.27)

Table 3.1 (continued)

Condition	Trial	Prestudy Mean (*SE*)	Poststudy Mean (*SE*)	Postdiction Mean (*SE*)
Global	1	11.50 (1.35)	7.04 (.85)	6.39 (.73)
	2	10.77 (1.10)	12.92 (1.25)	12.58 (1.31)
Combination w/Diff	1	10.36 (1.37)	5.92 (.86)	6.72 (.74)
	2	11.04 (1.12)	13.04 (1.28)	12.32 (1.34)
Combination	1	11.93 (1.32)	7.15 (.83)	6.19 (.72)
	2	11.70 (1.08)	13.11 (1.23)	12.56 (1.29)

Note. Predictions and Postdictions were between 0 and 36.

Follow up independent samples *t*-tests were conducted which compared the means of each condition relative to the other for the prestudy predictions, poststudy predictions, and postdictions for each trial. There were no significant differences between any of the conditions for prestudy and poststudy judgments made in trial 1. The majority of differences between conditions occurred in trial 2 where the global with difficulty condition participants were the most accurate in their judgments compared to participants in the other feedback conditions. The significant *t*-test values are reported in Table 3.2.

Table 3.2

Significant Independent t-tests for Predictions and Postdictions

Comparison	*t-value*	*p-value*	*Cohen's d*
Postdiction Trial 1			
Control vs. Global w/	*t*(52) = 2.26	.00	.62
Control vs. Global	*t*(50) = .32	.05	.09
Control vs. Combination	*t*(51) = .53	.04	.15
Item-level w/ vs. Item-level	*t*(52) = 3.62	.01	.99
Prestudy Trial 2			
Control vs. Global w/	*t*(52) = 2.16	.02	.59
Item-level w/ vs. Item-level	*t*(52) = 3.30	.01	.90
Global w/ vs. Combination w/	*t*(51) = -1.67	.00	-.46
Poststudy Trial 2			
Control vs. Global w/	*t*(52) = 1.79	.04	.49
Global w/ vs. Combination w/	*t*(51) = -1.87	.00	-.51
Combination w/ vs. Combination	*t*(50) = -.04	.03	-.01
Postdiction Trial 2			
Control vs. Global w/	*t*(52) = 2.99	.01	.81
Item-level w/ vs. Global w/	*t*(51) = 1.99	.00	.55
Global w/ vs. Combination w/	*t*(52) = -2.97	.00	-.82

Absolute accuracy of predictions and postdictions. The absolute accuracy for prestudy and poststudy judgments was analyzed by subtracting actual performance from predicted performance, with values closer to 0 indicating more accurate judgments and positive values indicating overestimation and negative values indicating underestimation. A 2 x 3 x 7 ANOVA was conducted with Trial (1 or 2) and Judgment type (prestudy prediction, poststudy prediction, or postdiction) as the within subjects factors and Condition as the between subjects factor. There was a significant main effect for Trial, $F(1, 176) = 218.70$, $p < .01$, $\eta_p^2 = .55$, because participants' accuracy decreased across trials (Trial 1 mean = -.01, *SE* = .01; Trial 2 mean = -.14, *SE* = .01). There was a significant main effect for Judgment type, $F(2, 175) = 40.13$, $p < .01$, $\eta_p^2 = .31$, because participants became less accurate across judgments (Prestudy prediction mean = -.03, *SE* = .01; Poststudy prediction mean = -.09, *SE* = .01; Postdiction mean = -.10, *SE* = .01). There was also a significant interaction between Trial and Judgment type, $F(2, 175) = 71.70$, $p < .01$, $\eta_p^2 = .45$, and Judgment type and Condition, $F(12, 352) = 2.04$, $p = .02$, $\eta_p^2 = .07$. There was no interaction between Trial and Condition, $F(6, 176) = 1.89$, $p = .09$.

Mean accuracy of JOLs. Delayed JOLs were collected for all 36 Chinese characters after the study phase in each trial. These judgments were based on how likely the participant thought they would be able to recall the English equivalent when presented with the Chinese character 10 min from that time with 0 being a judgment that they would be 0% likely and 100 being they are 100% certain they would recall the item on a recall test. A 2 x 3 x 7 ANOVA was conducted with Trial (1 or 2) and normative Item difficulty level (simple, moderate, complex) as the within subjects factors and

Condition as the between subjects factor. There was a significant main effect for Trial, $F(1, 176) = 3.40$, $p < .01$, $\eta_p^2 = .66$, because participants' JOLs increased across trials (Trial 1 mean = 32.09, $SE = 1.18$; Trial 2 mean = 50.93, $SE = 1.38$). There was a significant main effect for Item difficulty level, $F(2, 175) = 3.99$, $p < .01$, $\eta_p^2 = .82$, because JOLs decreased as item difficulty increased. There was also a significant interaction between Trial and normative Item difficulty level, $F(2, 175) = 19.63$, $p < .01$, $\eta_p^2 = .18$ (see Figure 3.3). However, no significant interaction between Trial and Condition was observed, $F(6, 176) = .74$, $p = .62$ or Item difficulty and Condition, $F(12, 352) = 1.25$, $p = .25$. The means and standard errors are reported in Table 3.3.

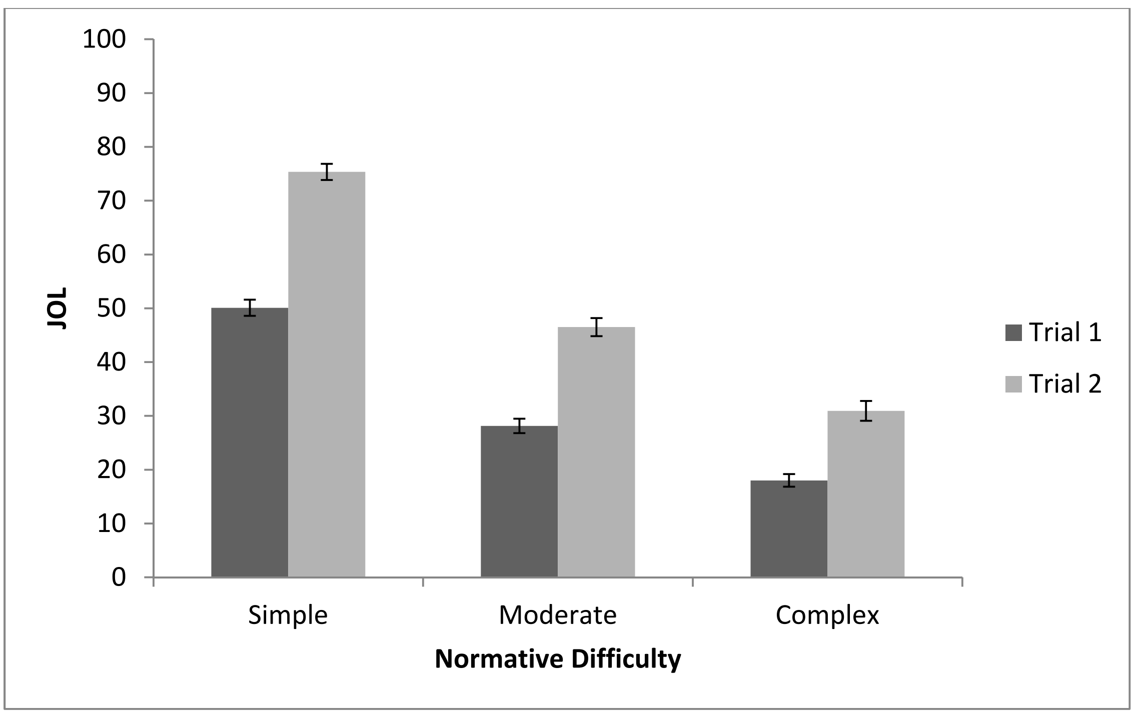

Figure 3.3: Mean JOLs for Trial 1 and 2 as a Function of Normative Item Difficulty.

Table 3.3

JOLs as a Function of Trial and Condition

	Trial 1	Trial 2
Condition	Mean (*SE*)	Mean (*SE*)
Control	32.03 (3.12)	49.60 (3.66)
Item-level w/Diff	29.52 (3.67)	48.63 (3.74)
Item-level	33.99 (3.12)	55.68 (3.66)
Global w/Diff	30.96 (3.01)	45.62 (3.53)
Global	33.19 (3.12)	53.61 (3.66)
Combination w/Diff	33.45 (3.18)	52.05 (3.74)
Combination	31.44 (3.06)	51.31 (3.60)

Note. Judgments were based on a 1 to 100 scale.

Absolute accuracy of JOLs. The absolute accuracy for JOLs was analyzed with values closer to 0 indicating more accurate judgments and positive values indicating overestimation and negative values indicating underestimation. A 2 x 3 x 7 ANOVA was conducted with Trial (1 or 2) and normative Item difficulty level (simple, moderate, complex) as the within subjects factors and Condition as the between subjects factor. There was a significant main effect for Trial, $F(1, 176) = 20.32$, $p < .01$, $\eta_p^2 = .99$, because participants' absolute JOL accuracy increased across trials (Trial 1 mean = .08, *SE* = .01; Trial 2 mean = .04, *SE* = .01). There was a significant main effect for Item difficulty level, $F(2, 175) = 13.10$, $p < .01$, $\eta_p^2 = .13$, because JOL accuracy generally

decreased as item difficulty increased (Simple mean = .04, *SE* = .01; Moderate mean = .08, *SE* = .01; Complex mean = .07, *SE* = .01). There was also a significant interaction between Trial and normative Item difficulty level, $F(2, 175) = 19.51$, $p < .01$, $\eta_p^2 = .18$ (see Figure 3.4). However, no significant interaction between Trial and Condition, $F(6, 176) = .80$, $p = .57$, or Item difficulty and Condition, $F(12, 352) = 1.37$, $p = .18$, was observed.

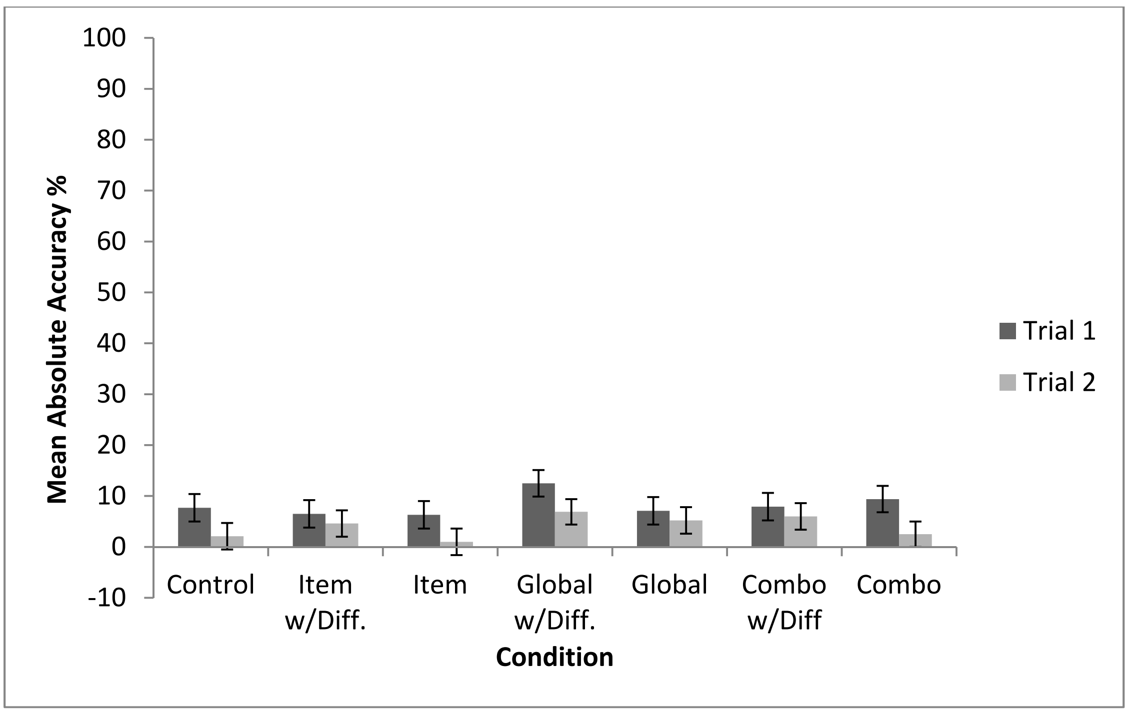

Figure 3.4: Absolute Accuracy of JOLs as a Function of Condition.

Relative Accuracy of JOLs. The relative accuracy of the delayed JOLs was calculated using gamma correlations (Nelson, 1984). These correlations compare the accuracy of one item relative to another instead of only the mean difference to allow examination of whether items given higher JOLs were more likely to be recalled than

items given lower JOLs. Values closer to 1 indicate that there was greater relationship between the participant's JOLs and their actual recall performance. A 2 x 7 ANOVA was conducted between Trial (1 or 2) and Condition. There was no significant main effect for Trial, $F(1, 171) = 3.03$, $p = .08$ and no significant interaction between Trial and Condition, $F(6, 171) = .73$, $p = .62$. The means and standard errors are reported in Table 3.4.

Table 3.4

Relative JOL Accuracy as a Function of Condition and Trial

	Trial 1	Trial 2
Condition	Mean (*SE*)	Mean (*SE*)
Control	.87 (.04)	.84 (.05)
Item-level w/Diff	.91 (.04)	.93 (.05)
Item-level	.84 (.04)	.90 (.05)
Global w/Diff	.86 (.04)	.86 (.05)
Global	.91 (.04)	.94 (.05)
Combination w/Diff	.88 (.04)	.93 (.05)
Combination	.89 (.04)	.96 (.05)

Note. Correlations closer to 1 indicate a stronger relationship between the JOL and actual recall performance.

Selection

Selection Order. The order in which items were selected for study within the 3 x 6 grid during the 45 s restudy phase was recorded for the first 18 item selections. Simple items were assigned a value of 1, moderate items a value of 2, and complex items a value of 3 for analysis (Metcalfe & Kornell, 2003). With these assigned values, the average difficulty level chosen by participants could be analyzed for the first item selected, the second, up through the eighteenth selection. Therefore if the first item selected had an average close to 1 then participants tended to select a simple item first and if the average was closer to 3 then they tended to select more a complex item. The means for these selections as a function of condition can be seen in Figure 3.5. An 18 x 7 ANOVA was conducted with Selection order (1 through 18) within subjects and Condition between. There was a significant main effect for Selection, $F(17, 134) = 2.90$, $p < .01$, $\eta_p^2 = .27$, but no significant interaction between Selection and Condition, $F(102, 834) = 1.05$, $p = .35$. The results indicate that participants either selected items at random for restudy or that something other than item difficulty drove their self-regulated selection decisions.

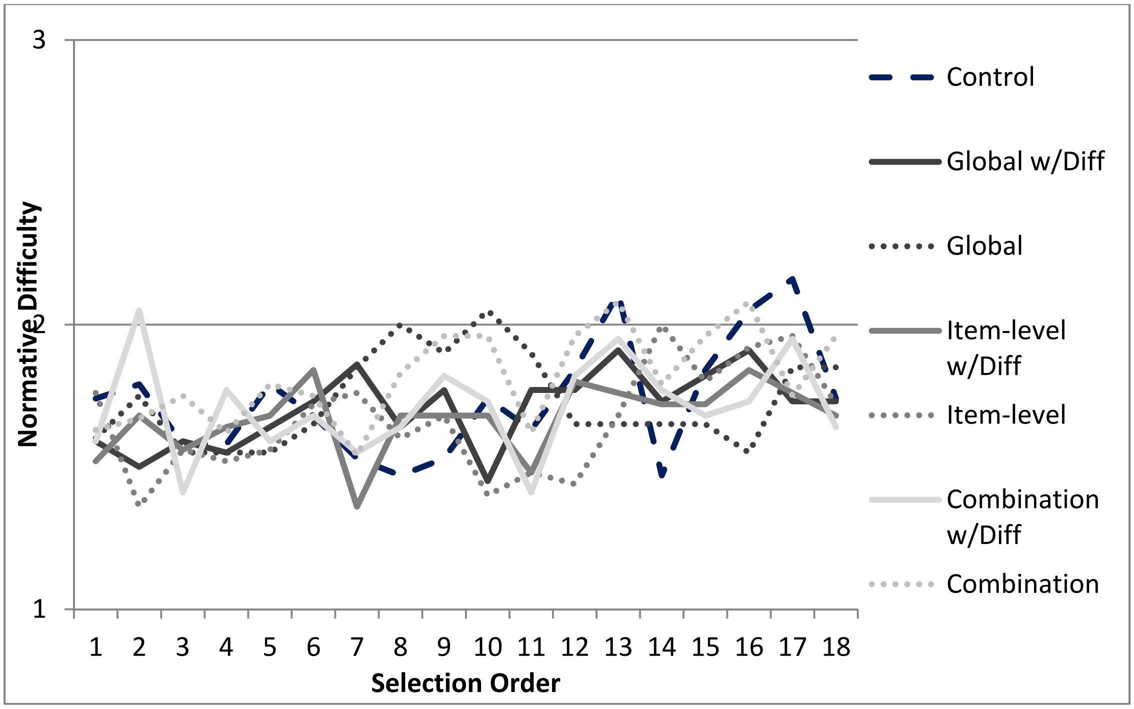

Figure 3.5: Selection Order as a Function of Item Difficulty.
Note. The average difficulty of the items selected (1 = Simple, 2 = Moderate, 3 = Complex) in the grid based on their selection order (1 through 18) and the condition.

Conditional selection accuracy. Given that the selection behaviors plotted in Figure 3.5 were constrained by which items were selected for restudy, we conducted additional analyses to examine whether participants' selections were optimal. Conditional probabilities for selection were calculated, which took into account whether the item had been recalled correctly on the trial 1 recall test. Items selected for restudy which were incorrectly recalled in the first trial were given a value of 1, indicating it was an optimal selection choice. Items selected for restudy which were correctly recalled in the first recall trial were given a value of -1, indicating it was not an optimal selection choice. Participants with averages closer to 1 were more accurate in their item selections. ANOVAs were conducted for the conditional selections of each level of Item difficulty and Condition. There was a significant effect for Simple items, $F(6, 176) = 2.24$, $p = .04$,

$\eta_p^2 = .27$, Moderate items, $F(6, 176) = 3.08$, $p < .01$, $\eta_p^2 = .31$, and Complex items, $F(6, 176) = 2.31$, $p = .04$, $\eta_p^2 = .27$. The mean for Simple items was .46 ($SE = .03$), the mean for Moderate items was .63 ($SE = .04$), and the mean for Complex items was .63 ($SE = .04$). The means for each condition collapsed across normative difficulty are presented in Figure 3.6.

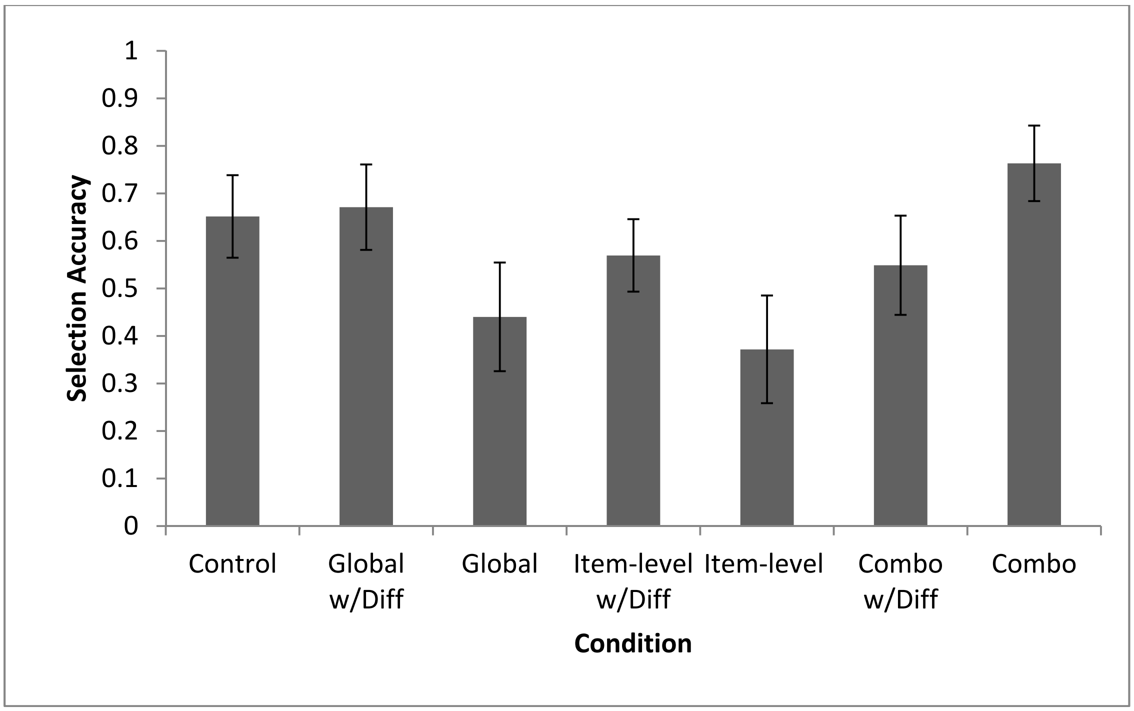

Figure 3.6: Conditional Selection Accuracy as a Function of Condition.
Note. Values closer to 1 indicate participants more optimally selected items for restudy.

Follow up independent samples *t*-tests were conducted which compared the conditional selection accuracy of each condition against one another. These yielded significant differences between the control compared to the item-level without difficulty, $t(50) = 1.79$, $p = .04$, $d = .50$, and the control compared to the global without difficulty conditions, $t(50) = 2.45$, $p = .03$, $d = .68$. Also the combination without difficulty

compared to the item-level with difficulty, $t(52) = -1.11$, $p = .02$, $d = -.31$, the combination without difficulty compared to the item-level without difficulty, $t(51) = -2.93$, $p < .01$, $d = -.81$, the combination without difficulty compared to the global without difficulty, $t(51) = -3.69$, $p < .01$, $d = -1.01$, and the combination without difficulty compared to the combination with difficulty, $t(50) = -2.11$, $p < .01$, $d = -.59$.

Study-time

The amount of time allocated to items which were selected for restudy was analyzed according to the normative difficulty of the items. Participants had a total of 45 s to allocate among the 18 items they chose for restudy. A 3 x 7 ANOVA was conducted with study-time analyzed as a function of Item Difficulty (simple, moderate, complex) and Condition. There was a significant main effect for Item Difficulty, $F(2, 175) = 49.27$, $p < .01$, $\eta_p^2 = .36$. This was expected because participants generally chose to study simple and moderately complex items for restudy instead of complex items. For this reason participants spent the majority of their time on the second trial studying the simple items ($M = 17.91$, $SE = .67$), followed by the moderate items ($M = 14.77$, $SE = .44$), and the least amount of time was spent on complex items ($M = 7.86$, $SE = .58$). There was no significant interaction between Item Difficulty and Condition, $F(12, 352) = .95$, $p = .50$.

The proportion of study time allocated to items within each difficulty level was also analyzed as a function of condition. A 3 x 7 ANOVA was conducted. There was a significant main effect for Item difficulty, $F(2, 175) = 54.15$, $p < .01$, $\eta_p^2 = .38$. This was also expected because there were generally more simple items in the grid than complex items; thus, participants should have spent proportionally more time on the easier items.

There was no significant interaction between Item difficulty and Condition, $F(12, 352) = .06$, $p = .57$. The proportions of time spent on items within each level of difficulty are reported in Table 3.5.

Table 3.5

Proportion of Time Spent on Items as a Function of Condition

	Simple	Moderate	Complex
Condition	Mean (*SE*)	Mean (*SE*)	Mean (*SE*)
Control	.46 (.04)	.35 (.03)	.19 (.04)
Item-level w/	.40 (.04)	.40 (.03)	.20 (.04)
Item-level	.44 (.04)	.36 (.03)	.21 (.04)
Global w/	.49 (.04)	.35 (.02)	.16 (.04)
Global	.49 (.04)	.36 (.03)	.15 (.04)
Combination w/	.44 (.04)	.33 (.03)	.23 (.04)
Combination	.39 (.04)	.39 (.02)	.22 (.04)

Recall Performance

Recall performance was recorded in both trials and analyzed based on normative difficulty level. All responses were additionally handscored during analysis in case any misspellings or typing errors were missed by the computerized scoring process to insure proper analysis of recall. A 2 x 3 x 7 ANOVA was conducted with Trial (1 or 2), Difficulty level (simple, moderate, or complex), and Condition. There were significant

main effects for Trial, $F(1, 176) = 1.00$, $p < .01$, $\eta_p^2 = .85$, and Difficulty level, $F(2, 175) = 603.93$, $p < .01$, $\eta_p^2 = .87$, because recall performance decreased as difficulty level increased, despite the fact that recall increased across trials for all three difficulty levels. There was a significant interaction between Trial and Difficulty level, $F(2, 175) = 66.25$, $p < .01$, $\eta_p^2 = .43$. The interaction was driven by the fact that recall performance improved across trials as participants were exposed to the same material in the second trial and participants recalled more of the simple items than the more complex items across both trials. The means and standard errors for Difficulty level and Trial are presented in Figure 3.7.

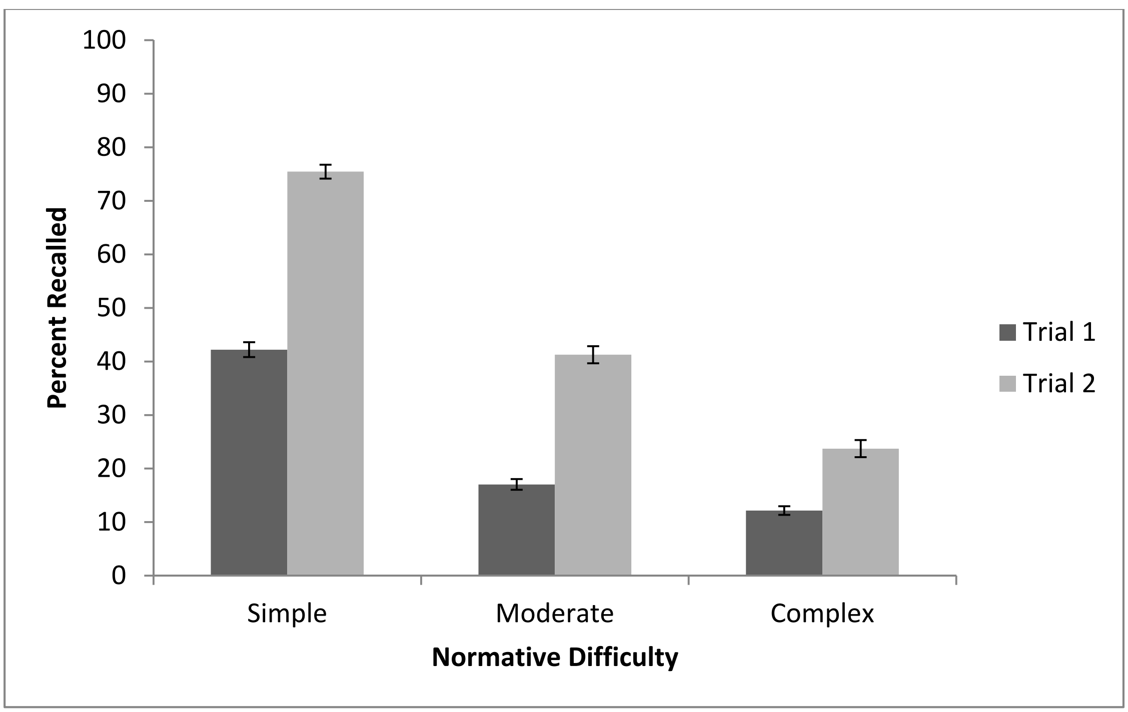

Figure 3.7: Recall Accuracy in Each Trial as a Function of Item Difficulty.

There were also significant interactions between Trial and Condition, $F(6, 176) = 2.23$, $p = .04$, $\eta_p^2 = .07$, and Difficulty level and Condition, $F(12, 352) = 1.81$, $p = .05$, $\eta_p^2 = .06$. The overall mean recall, collapsing across conditions and normative difficulty levels, was .35 (*SE* = .01). The means for each condition and trial are reported in Table 3.6.

Table 3.6

Recall Accuracy as a Function of Trial and Condition

	Trial 1	Trial 2
Condition	Mean (*SE*)	Mean (SE)
Control	.24 (.02)	.48 (.03)
Item-level w/Diff	.23 (.02)	.44 (.03)
Item-level	.27 (.02)	.56 (.03)
Global w/Diff	.19 (.02)	.39 (.03)
Global	.26 (.02)	.48 (.03)
Combination w/Diff	.26 (.02)	.46 (.03)
Combination	.22 (.02)	.49 (.03)

Follow up independent samples *t*-tests were conducted which compared the recall performance of each condition against one another based on the normative item difficulty and the trial. These yielded significant differences for simple items in trial 1 for the item-level without difficulty compared to the global with difficulty condition, $t(52) = 2.79$,

p = .02, *d* = .76, and for the comparison between the global with difficulty and the global without difficulty *t*(52) = -2.16, *p* = .02, *d* = -.59. There were significant differences between conditions for moderate items in trial 1 with the control compared to the item-level without difficulty, *t*(52) = .94, *p* = .03, *d* = .26, the control compared to the combination without difficulty, *t*(51) = 1.65, *p* = .03, *d* = .45, and the combination with difficulty compared to the combination without difficulty, *t*(50) = 2.23, *p* = .04, *d* = .62 . There were also significant differences between conditions for complex items in trial 2 with the control compared to the item-level without difficulty, *t*(52) = 1.82, *p* < .01, *d* = .50, the item-level without difficulty and the global with difficulty, *t*(52) = 2.94, *p* = .02, *d* = .80, and item-level without difficulty compared to the combination without difficulty, *t*(53) = -2.57, *p* = .01 *d* = -.69. There were no significant differences between conditions for complex items in trial 1, simple items in trial 2, and moderate items in trial 2.

CHAPTER IV

DISCUSSION

The main focus of this experiment was to explore how the placement of feedback influenced self-regulated learning. In particular, we were interested in feedback's influence over the metacognitive awareness of participants, the items participants selected for restudy, the order in which participants studied items, the amount of time spent on those items, and their subsequent recall performance. We manipulated when participants received feedback (global, item-level, or combination) and whether or not they received information about how difficult the items were. We expected conditions which had more proximal and informative feedback to be more beneficial to the learner. The results provided mixed support to this generalized hypothesis indicating that feedback may have a more intricate and complex relationship to self-regulated learning which warrants further investigation.

We included information about normative item difficulty in half of the feedback conditions in order to observe whether this information would affect reselection behaviors. These behaviors included which items participants chose to study, in which order, and how long they studied each item before moving on to the next item (Metcalfe & Kornell, 2005). Participants generally chose to study simple and moderately complex items for restudy instead of complex items. This supports the RPL model as

people selected easier items (Metcalfe, 2002). The influence of normative item difficulty information varied widely with the type of feedback participants received. Selection accuracy varied as a function of condition with the item-level without difficulty information condition having the lowest overall mean and the global without difficulty information having the second lowest overall mean. This may indicate that the item difficulty information was helpful to participants when making their decisions on which items to study. Although not significantly different, the combination without difficulty information condition had the highest selection accuracy average. These results may indicate that with specific feedback at the global and item-level, difficulty information can serve to confuse the participants when making their selection choices. In turn, this may increase the participants' cognitive load and decrease learning and performance (Paas & Van Merriënboer, 1994).

The order in which items were selected for study in the second trial was analyzed to observe whether participants selected items based on their difficulty and whether this would be altered as a function of the condition to which they were assigned. Participants appeared to select the 18 items at random, as there was no difference in normative difficulty level across the selection process. This generally does not support the RPL model where participants would be expected to select the easiest items in the grid first and then move on to the more difficult items as they progressed through the study session (Metcalfe, 2002). It also does not support the DRM which states participants would concentrate on the more difficult items first before moving on to the easier items (Dunlosky & Hertzog, 1998). Selection within the grid did not reveal any clear cut RPL or DRM selection behaviors; instead selection appeared to be random. This pattern

suggests there was something other than the normative difficulty of the items driving the selection order for restudy. Participants may have simply chosen in a left-to-right selection pattern which would appear random as items were randomly placed in the grid. This would offer support to the ABR model which incorporates habitual responding, such as reading order, as one of the variables which can influence self-regulated study decisions (Ariel et al., 2009; Ariel et al., 2011; Dunlosky & Ariel, 2011).

Participants spent the majority of their time on the second trial studying the simple items, followed by the moderate items, and the least amount of time was spent on complex items. Similar results were observed when the amount of study time was analyzed proportionally to the other items in the grid. This supports the RPL model as people spent more time on easier items (Metcalfe, 2002). Participants allocated proportionately the least amount their total study time to the complex items. Thus results did not offer any support for the DRM's prediction that people choose to allocate more of their total study time to the difficult items (Dunlosky & Thiede, 1998). The results do support the predictions of the ABR model since the model incorporates item difficulty as an influence over the agendas individuals form while studying (Ariel et al., 2009).

Recall performance may explain why participants selected easier and moderate items for restudy. The overall recall was only 35% so perhaps participants generally did not learn enough of the simple and moderate items to be able to move on to more complex items. Recall accuracy also varied as a function of condition with the item-level with difficultly information having the lowest accuracy and the global condition without difficulty information having the highest at accuracy. This difference may be because feedback at the item-level with additional difficulty information may be too complex for

participants to be able to track as compared to a general estimation of how many of the items they missed.

We predicted that providing participants with more information in the feedback process would benefit the learner by giving them more metacognitive monitoring. It was expected that participants would have better metacognitive monitoring, as measured by relative JOL accuracy, in conditions where they received more proximal and informative feedback. The results showed that the relative accuracy of JOLs was not significantly different as a function of condition. Although metacognitive monitoring did not differ as a function of condition, this calculation allowed us to observe that participants were generally metacognitively aware of which items they would and would not recall on a subsequent recall test, with an average accuracy of approximately 90%. This high level of metacognitive monitoring could explain the nonsignificant results. As participants displayed a high level of metacognitive monitoring with their JOLs, any differences between conditions may have been too minor to affect the overall accuracy.

This high amount of metacognitive monitoring may be the result of the stimuli used for this experiment. Participants made their predictions as to whether or not they would be able to correctly recall an item on a subsequent recall test while seeing the Chinese character without the English translation. Participants tend to be more accurate in their JOLs when they are only presented with half of the stimulus pair (Dunlosky & Nelson, 1997). This may have led to an increase in confidence as to whether they would be able recall an item or not. The JOLs were also collected after a delayed amount of time which has been shown to yield more accurate JOLs than those collected immediately after learning an item (Nelson & Dunlosky, 1991). The format, the timing,

and the type of stimuli used for the JOLs may have contributed to the amount of metacognitive monitoring exhibited across all conditions thus this may have produced a ceiling effect for metacognitive decisions.

It was expected that immediate item-level feedback would be more beneficial than global feedback immediately following a recall test (Kulik & Kulik, 1988). We also expected that giving additional information as to the complexity of the item would aid in the learning process by allowing a person to make a generalization about which level of complexity they need to study the most so that they do not need to memorize the exact items they got incorrect. These predictions implied that participants would be able to accurately monitor their performance without causing too much cognitive load.

Cognitive load refers to strain on cognitive processes due to the complexity of the task or the limitation of human working memory which can interfere with the learning of new materials (Paas & Van Merriënboer, 1994). Cognitive load theory describes three types of cognitive load: germane, intrinsic, and extraneous load. Germane load refers to the working memory resources necessary to complete the task, such as strategies and organization. Intrinsic load is caused by the internal complexity of the task. Extraneous load is caused by extra information which is not necessary for the task, such as contradictory or redundant instructions (Sweller, van Merrienboer, & Paas, 1998). Any intrinsic and extraneous load added to a task can take away from the germane load which is necessary to perform the task optimally. Thus, if the task itself is overly complex or if the instructions are confusing to the participant, then there would be an increase in cognitive load which could negatively affect performance.

Cognitive load may have influenced the results of the present experiment. The task itself may have been too complex for the participants and caused an increase in intrinsic load. This would include the stimuli used in the experiment as Chinese is a character based language which may be difficult for a native English speaker to learn. Also the fact that the participants were asked to learn Chinese under timed conditions using a computerized program may have increased the complexity if this format did not align with how an individual would normally approach learning a novel language. As there were few differences between conditions, the feedback may have added extraneous load instead of easing the germane cognitive load required for the task. The feedback was intended to give external information about a participant's performance to relieve the cognitive strain of tracking whether items were correct or incorrect as well as which type of item complexity to focus on. If participants were unable to use the feedback to their advantage, then the feedback would have served as superfluous information to add to an already complex task and thus could have increased the participants' overall cognitive load.

Participants in the item-level only feedback conditions may have found it difficult to track which items they got correct or incorrect and may not have been able to properly aggregate their performance across trials. Although participants in the global feedback conditions should have had a reduced cognitive load by not having to aggregate their performance, they may not have been able to determine which specific items they got correct or incorrect. Only participants in the combination conditions which received both item-level and global feedback should have had the benefit of a reduced cognitive load by not having to aggregate their performance as well as the benefit of knowing which items

they specifically got correct or incorrect. Participants in the combination condition, particularly in the condition with difficulty information, should have had the lowest cognitive load and thus been the most aware of which level of difficulty they needed to focus on as well as the precise items to restudy. This did not hold true for the present experiment. Selections were the most accurate for the combination condition without difficulty information and recall performance was similar across conditions. This may indicate that the difficulty information was not enough to reduce the cognitive demands of the task or it actually served to increase cognitive load.

As with most experiments, there are some limitations. Although previous research has shown elaborated feedback to be able to help reduce cognitive load (Moreno, 2004) the difficulty of the task may have been too much for participants. There may have been too much information and the material may have been too novel for the participants, resulting in too large of a cognitive load for individuals to both learn the items and process the feedback effectively. The multimedia approach to this experiment may have also increased participants' cognitive load because participants had to use an interactive computer program with which they had no prior experience. Novel multimedia use has been shown to increase cognitive demand and therefore may have hindered performance since participants were not given a chance to practice with the program before the experiment began (Mayer & Moreno, 2003). It remains unclear as to how much feedback an individual needs to promote optimal self-regulated learning behaviors. This may be due to individual differences or situational differences which could not be teased apart with the present experiment.

Another possible limitation is that we used Chinese stimuli which may not directly translate to other learning tasks. Although the Chinese characters may be different from the typical material learned, this fact may also be one of the strengths of the experiment since participants should not have been able to guess the meaning of the novel material thereby allowing us to observe the true process of learning and not a person's ability to make educated guesses. While the Chinese stimuli may be strong experimentally, they still may not provide a true reflection of how people self-regulate their learning. A person's behavior in a computerized task, in which good or poor performance does not affect them, may not be the same as their behavior in a classroom or employment related environment. Even with these limitations, this experiment should provide direction for the specificity of subsequent experiments.

By better understanding how people allocate their time during self-regulated learning, training strategies can be developed to help people be more optimal in their efforts and increase their learning productivity. This is especially important as society is placing an increased focus on online and distance learning programs where individuals have to regulate learning on their own. Research into how to optimally allocate study time is vital as self-regulation is becoming more necessary. It is important that people are metacognitively aware of their study behavior so they can better control their behavior (Nelson & Narens, 1990). This experiment provides insight into how feedback can aid in self-regulated learning and optimize performance.

The present experiment investigated the placement of feedback and its influence in a self-regulated learning task. Understanding where feedback should be placed in order to be the most beneficial to the learner is important for the enhancement of the

learning process. The results from this experiment provide information for further investigation into the feedback process. Once the optimal placement has been established then exploration into the type of feedback information to provide as well as the format of the presentation could be explored with the ultimate goal of improving learning performance. By providing optimal feedback a person should be able to improve their metacognitive control during the learning process, which should in turn enhance their learning.

APPENDICES

APPENDIX A

UHSC Form

UAHuntsville
UNIVERSITY OF ALABAMA IN HUNTSVILLE

Nicholas Jones
332B Morton Hall
Phone: 256 824 2338
Fax: 256.824 2387
Email: irb@uah.edu

Jodi Price, Ph.D.
Department of Psychology
UAHuntsville
Huntsville, AL 35899

January 22, 2011

Dear Dr. Price,

As chair of the IRB Human Subjects Committee, I have reviewed your proposal, *Does feedback about performance influence novice younger and older adults' approach to learning Chinese characters?*, and have found it meets the necessary criteria for expedited review according to 45 CFR 46. I have approved this proposal, and you may commence your research. Please note that this approval is good for one year from the date on this letter. If data collection continues past this period, a renewal application must be filed with the IRB.

Please contact me if you have any questions.

Sincerely,

Dr. Nicholas Jones
Chair, UHSC

OFFICE OF THE VICE PRESIDENT FOR RESEARCH
Von Braun Research Hall M-17 Huntsville, AL 35899 T 256.824.6100 F 256.824.6783

APPENDIX B

PERSONAL DATA SHEET

In order to better understand the results of our study, we need to gather some information about you and your background. This information is for research purposes only, and will be kept strictly confidential. You may note that we do not ask for your name on this form. Please respond to the following items completely.

1. Gender: 1- [ ] Male
   2- [ ] Female

2. Age: _______

3. Date of Birth: _____/ _____/ ______
   Month Day Year

4. What is your native language?
   1- [ ] English
   2- [ ] Other (please specify) _____________

4a. Do you consider yourself to be Hispanic or Latino? (See definition below). Mark ONE. ***Hispanic or Latino:*** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture, regardless of race.
   1- [ ] Hispanic or Latino
   2- [ ] Not Hispanic or Latino

4b. What race do you consider yourself to be? (Mark **all** that apply)
   1- [ ] **American Indian or Alaska Native.** A person having origins in North, Central, *or* South America, and who maintains tribal affiliation or community attachment.
   2- [ ] **Native Hawaiian or Other Pacific Islander.** A person having origins in Hawaii, Guam, Samoa, or *other* Pacific Islands.
   3- [ ] **Asian**. A person having origins in the Far East, Southeast Asia, or the Indian *subcontinent*, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippines, and Vietnam.
   4- [ ] **White.** A person having origins in Europe, the Middle East, or North Africa.
   5- [ ] **Black or African American.** A person having origins in any of the black racial groups of Africa.

5. What is your **current** marital status?

1- [ ] Married	4- [ ] Divorced
2- [ ] Single	5- [ ] Separated
3- [ ] Living with Partner	6- [ ] Widowed

6. Do you have children?
1- [ ] Yes (please indicate how many) __________
2- [ ] No

6b. Do you have grandchildren?
1- [ ] Yes (please indicate how many) __________
2- [ ] No

7. What type of dwelling do you live in (Mark ONE)?
1- [ ] Single Family House
2- [ ] Duplex/Townhouse
3- [ ] Apartment or Condominium
4- [ ] Campus Housing (dorm, fraternity, sorority, etc)
5- [ ] Congregate or Senior Care Facility
6- [ ] Other (please specify) ________________

8. What is the **HIGHEST** grade or level of school **COMPLETED** or the **HIGHEST** degree you have received? (Mark **ONE** box ONLY)
0- [ ] Never Attended/ Kindergarten Only
1- [ ] 1st Grade
2- [ ] 2nd Grade
3- [ ] 3rd Grade
4- [ ] 4th Grade
5- [ ] 5th Grade
6- [ ] 6th Grade
7- [ ] 7th Grade
8- [ ] 8th Grade
9- [ ] 9th Grade
10-[ ] 10th Grade
11-[ ] 11th Grade
12-[ ] 12th Grade, **No Diploma**
13-[ ] **High School Graduate** – high school DIPLOMA or the equivalent (Example: GED)
14- [ ] Some college credit, but less than one year
15- [ ] 1 or more years of college, no degree
16- [ ] Associate Degree: Occupational, Technical, or Vocational Program
17- [ ] Associate Degree: Academic Program
18- [ ] Bachelor's Degree (Example: BA, AB, BS, BBA)
19- [ ] Master's Degree (Example: MA, MS, MEng, Med, MBA)
20- [ ] Professional School Degree (Example: MD, DDS, DVM, JD)
21- [ ] Doctoral Degree (Example: PhD, EdD)

9. What is your PRIMARY occupational status? (Mark **ONE** box ONLY)
1- [ ] Full time Job ! Skip to 9a
2- [ ] Part time Job ! Skip to 9a
3- [ ] Full time Homemaker ! Skip to 10

4- [ ] Student ! Skip to 9c
5- [ ] Self-Employed ! Skip to 9a
6- [ ] Retired ! Skip to 9b
7- [ ] Not Employed ! Skip to 10

9a. Briefly describe your current job (*If you have more than one job, describe the one you worked the most hours*): ____
____

9b. Briefly describe your primary occupation prior to retirement (*If you had more than one job, please describe the one you worked the most hours):*
____
____

9c. What is your major? ____

10. In general would you consider your overall health to be (Mark ONE):
1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor

11. Compared to a perfect state of health, I believe my overall health to be (Mark ONE):
1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor

12. Compared to other people my age, I believe my overall health to be (Mark ONE):
1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor

13. Compared to other people my age, I believe my eye-sight, ***WITHOUT* the aid of corrective lenses/ contacts or glasses,** to be (Mark ONE):
1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor

13a. Compared to other people my age, I believe my eyesight, ***WITH* the aid of corrective lenses/ contacts or glasses,** to be (Mark ONE):

1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor
6- [ ] N/A

14. Compared to other people my age, I believe that my hearing, ***WITHOUT* the help of devices such as hearing aids,** to be (Mark ONE):

1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor

14a. Compared to other people my age, I believe that my hearing, ***WITH* the help of devices such as hearing aids,** to be (Mark ONE):

1- [ ] Excellent
2- [ ] Very Good
3- [ ] Good
4- [ ] Fair
5- [ ] Poor
6- [ ] N/A

15. Are you currently experiencing any problems with your memory or your ability to pay attention that have been diagnosed by a doctor?

1- [ ] Yes (please explain) ____
2- [ ] No

16. Are you currently seeing a doctor for any medical problems?

1- [ ] Yes (please explain) ____
2- [ ] No

17. Are you currently taking any medications that make you drowsy, and in turn, make it difficult for you to concentrate?

1- [ ] Yes (please explain) ____
2- [ ] No

THANK YOU!!!

APPENDIX C

Chinese Character Stimuli

Simple			Moderate			Complex		
Character	**English**	**Stroke #**	**Character**	**English**	**Stroke #**	**Character**	**English**	**Stroke #**
人	**PERSON**	**2**	们	**DOOR**	**5**	家	**HOME**	**10**
匕	**SPOON**	**2**	北	**NORTH**	**5**	桃	**PEACH**	**10**
刀	**KNIFE**	**2**	好	**GOOD**	**6**	贼	**THIEF**	**10**
山	**MOUNTAIN**	**3**	在	**NOW**	**6**	黄	**YELLOW**	**11**
小	**SMALL**	**3**	同	**SAME**	**6**	猜	**GUESS**	**11**
中	**MIDDLE**	**4**	合	**CLOSE**	**6**	剩	**REMAIN**	**12**
文	**LANGUAGE**	**4**	光	**LIGHT**	**6**	裂	**CRACK**	**12**
云	**CLOUD**	**4**	各	**SEPARATE**	**6**	圜	**CIRCLE**	**13**
天	**SKY**	**4**	学	**STUDY**	**8**	解	**LOOSEN**	**13**
不	**NO**	**4**	性	**NATURE**	**8**	搬	**MOVE**	**13**
日	**SUN**	**4**	说	**EXPLAIN**	**9**	满	**FULL**	**13**
王	**KING**	**4**	活	**ALIVE**	**9**	幕	**GRAVE**	**14**

APPENDIX D

Post-Task Questionnaire

(Combination with Difficulty Information)

In the last task you were asked to try to learn Chinese-English vocabulary pairs. So that we may gain a better understanding of how you approached learning the vocabulary pairs, please answer the following questions.

1. In the first part of the experiment you were asked to rate how easy or difficult you thought each Chinese word would be to learn. Did your initial impressions of item difficulty change once you began studying items (i.e., once you were able to see what each Chinese word means in English)? YES NO

2. If yes, how did your impressions change (e.g., did the Chinese words seem easier or more difficult once you were able to see the English translations)?

____

____

3. Is there anything that made certain Chinese characters easier to learn than others? If so, please explain.

____

____

4. Did you try to form or see an image in the Chinese characters that would help you remember the English equivalent (e.g., thinking that the Chinese character for *forest* looks like two trees)? YES NO

5. If you did use imagery to memorize the items, do you believe it helped you? YES NO

6. Is there any other strategy (other than imagery) you would recommend to others to help them memorize the items? YES NO

7. If yes, what?

____

____

8. On average, how much time do you think you would need to learn a single word pair?

____

9. Other than having more time, is there any other factor that you think influenced your ability to learn the vocabulary pairs? YES NO

10. If yes, what?

____

____

11. You were asked to make a judgment between 0 and 100 percent of how likely you would be able to recall each item on a later test. Do you think your judgments were accurate for *Trial 1*? YES NO

12. If no, why?

____

____

13. Do you think your judgments were accurate for *Trial 2*? YES NO

14. If no, why?

____

____

15. In the recall test, did you recall as many words as you had hoped in *Trial 1*? YES NO

16. If no, why?

____

____

17. In the recall test, did you recall as many words as you had hoped in *Trial 2*? YES NO

18. If no, why?

____

____

19. For *Trial 2*, if you had to summarize your strategy for <u>reselecting</u> items, what would it be?

____

____

20. Did you find the feedback given after <u>each</u> tested word about whether your response was correct or incorrect to be helpful? YES NO

21. If yes, how?

____

____

22. Did you find the feedback given at the <u>end</u> of the recall test about how many of the words you got correct to be helpful? YES NO

23. If yes, how?

___

___

24. Which type of feedback did you find to be more helpful? (Circle one)
    a. Feedback given after each recalled word
    b. Feedback given at the end of the recall test

25. Did you find the information given about item difficulty (i.e. whether the item was simple, moderate, or complex) to be helpful? YES NO

26. If yes, how?

___

___

27. Did the information about item difficulty influence how you chose to <u>reselect</u> items for study in *Trial 2*? YES NO

28. If yes, how?

___

___

29. How important was it to you to perform well on the recall tests? Circle (1 = not at all important and 7 = very important).

1 2 3 4 5 6 7

30. How enjoyable did you find this task to be? (1 = not enjoyable at all and 7 = very enjoyable).

1 2 3 4 5 6 7

31. We would like to examine whether strategy use differs as a function of students' grade point average (GPA). Please check in which group your overall GPA falls:

___ 3.5 – 4.0	___ 1.5 – 1.9
___ 3.0 – 3.4	___ 1.0 – 1.4
___ 2.5 – 2.9	___ Less than 1.0
___ 2.0 – 2.4	___ I'd rather not say

Please notify the experimenter that you are finished. Thank you!

REFERENCES

Ariel, R., Al-Harthy, I. S., Was, C. A., & Dunlosky, J. (2011). Habitual reading biases in the allocation of study time. *Psychonomic Bulletin & Review*, *18*(5), 1015-1021. doi:10.3758/s13423-011-0128-3

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 133,* 432-447. doi:10.1037/a0015928

Bangert-Drowns, R. L., Kulik, C .C., Kulik, J. A., & Morgan, T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213-38. doi: 10.3102/00346543061002213

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*, 211-241. doi:10.1037/a0022106

Berry, J. M., & West, R. L. (1993). Cognitive self-efficacy in relation to personal mastery and goal setting across the life span. *International Journal Of Behavioral Development*, *16*(2), 351-379.

Butler, A. C., Karpicke, J. D., & Roediger, H. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918-928. doi:10.1037/0278-7393.34.4.918

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245. doi:10.3102/00346543065 003245

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491-1494. doi:10.1037/0278-7393.27.6.1491

Chiviacowsky, S., & Wulf, G. (2007). Feedback after good trials enhances learning. *Research Quarterly For Exercise & Sport*, *78*(2), 40-47. http://www.aahperd.org/ rc/publications/rqes/

Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, *53*(4), 533-548. http://thepsychologicalrecord.siuc.edu/

Dunlosky, J., & Ariel, R. (2011). The influence of agenda-based and habitual processes on item selection during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 899-912. doi:10.1037/a0023064

Dunlosky, J., Ariel, R., & Thiede, K. W. (2011). Agenda-based regulation of study-time allocation. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of Remembering and Metacognition: Essays in honour of Bruce Whittlesea* (pp. 182 –198). Basingstoke: Palgrave MacMillan

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 249-275). Mahwah, NJ: Erlbaum

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1180-1191. doi:10.1037/0278-7393.27.5.1180

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language, 36,* 545-565. doi:10.1006/jmla.1996.2476

Dunlosky, J., & Thiede, K. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*(1), 37-56. doi:10.1016/S0001-6918(97)00051-6.

Dunlosky, J. & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier materials effect in control of study. *Memory and Cognition, 32*, 779-788. doi: 10.3758/BF03195868

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349-370. doi:10.1037/0096-3445.126.4.349

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 609-622. doi:10.1037/0278-7393.32.3.609

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79-97. doi:10.2307/1170349

Lachman, M. E., Bandura, M., Weaver, S. L., & Elliott, E. (1995). Assessing memory control beliefs: The Memory Controllability Inventory. *Aging & Cognition*, *2*(1), 67-84. doi:10.1080/13825589508256589

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. doi:10.1207/S15326985EP3801_6

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? J*ournal of Experimental Psychology: General, 131*(3), 349-363. doi:10.1037/0096-3445.131.3.349

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163. doi:10.1111/j.1467-8721.2009.01628.x

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174-179. doi:10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, *132*(4), 530-542. doi:10.1037/0096-3445.132.4.530

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463-477. doi:10.1016/j.jml.2004.12.001

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225-229. doi: 10.3758/BF03194056

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*(1-2), 99-113. doi:10.1023/B:TRUC.0000021811.66966.1d

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109-133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The 'delayed-JOL effect.'. *Psychological Science*, *2*(4), 267-270. doi:10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*(4), 207-213. doi:10.1111/j.1467-9280.1994.tb00502.x

Nelson, T., & Leonesio, R. (1988). Allocation of self-paced study time and the 'labor-in-vain effect.'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676-686. doi:10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). New York: Academic Press

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. doi: 10.1080/03075070600572090

Paas, F. C., & Van Merriënboer, J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351-371. doi:10.1007/BF02213420

Pelegrina, S., Bajo, M., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory*, *8*(6), 377-392. doi:10.1080/09658210050156831

Price, J., Hertzog, C., & Dunlosky, J. (2010). Self-regulated learning in younger and older adults: Does aging affect metacognitive control? *Aging, Neuropsychology, and Cognition, 17,* 329-359. doi:10.1080/13825580903287941

Price, J., & Murray, R. (under revision). Self-regulated learning of Chinese characters: Do younger, middle-aged, and older adults' attempts at regulation differ? *Psychology & Aging*

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 305-320. doi:10.1037/0278-7393.33.2.305

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763-776. doi:10.1037/0012-1649.27.5.763

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189. doi:10.3102/0034654307313795

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1236-1242. doi:10.1037/a0023548

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204-221. doi:10.1037/0278-7393.26.1.204

Sweller, J., van Merrienboer, J. G., & Paas, F. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296. doi:10.1023/A:1022193728205

van der Kleij, F. M., Eggen, T. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*(1), 263-272. doi:10.1016/j.compedu.2011.07.020

West, R. L., Welch, D. C., & Thorn, R. M. (2001). Effects of goal-setting and feedback on memory performance and beliefs among older and younger adults. *Psychology & Aging*, *16*(2), 240. doi:10.1037//0882-7974.16.2.240

Wulf, G., Chiviacowsky, S., & Lewthwaite, R. (2010). Normative feedback effects on learning a timing task. *Research Quarterly for Exercise and Sport*, *81*(4), 425-431. http://www.aahperd.org/rc/publications/rqes/